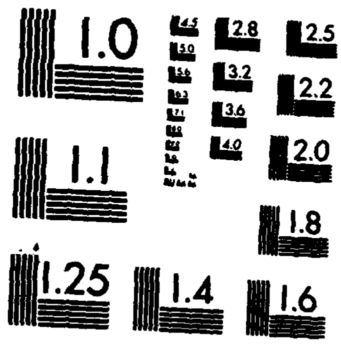
AD-A139 970 APPROXIMATIONS FOR EXPECTED VALUES OF NORMAL ORDER STATISTICS WITH AN APP..(U) STANFORD UNIV CA DEPT OF STATISTICS N FOTOPOULOS ET AL. 06 MAR 84 TR-342 1/1

UNCLASSIFIED N00014-76-C-0475 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

AD A139970

APPROXIMATIONS FOR EXPECTED VALUES [illegible]

[illegible]ITH AN APPLICATION TO [illegible]

BY

N. FOTOPOULO[illegible]

TECHNICAL REPORT NO. [illegible]

MARCH 6, 198[illegible]

PREPARED UNDER CONT[illegible]

N00014-[illegible]-C-0475 [illegible]

FOR THE OFFICE OF NAVAL [illegible]

Reproduction [illegible]

for any purpose [illegible]

Approved for public [illegible]

DEPARTMENT [illegible]

[illegible]

[illegible]

DTIC FILE COPY

APR 11 1984

84 04 09 048

APPROXIMATIONS FOR EXPECTED VALUES OF NORMAL ORDER STATISTICS
WITH AN APPLICATION TO GOODNESS-OF-FIT

BY

N. FOTOPOULOS, J. LESLIE and M. STEPHENS

TECHNICAL REPORT NO. 342
MARCH 6, 1984

Prepared Under Contract
N00014-76-C-0475 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

1. INTRODUCTION

1.1 Correlation Statistics

In this article results are given on the large-sample behaviour of normal order statistics. These are then used to demonstrate that two correlation statistics for testing normality, introduced at about the same time by Shapiro and Francia (1972) and by De Wet and Venter (1972), have the same asymptotic distribution, given and tabulated by De Wet and Venter (1972).

A correlation statistic for goodness-of-fit is one based on the usual correlation coefficient for pairs of random variables.

The definition is extended to give a "correlation" between order statistics $X_{1n} \leq X_{2n} \leq \cdots \leq X_{nn}$ of a random sample of size n and a suitable set of constants t_{in}, $i=1,2,\ldots,n$. The correlation statistic $r_n(X,t)$ is then defined by

$$r_n(X,t) = \frac{\sum_i (X_{in}-\bar{X})(t_{in}-\bar{t})}{\{\sum_i (X_{in}-\bar{X})^2 \cdot \sum_i (t_{in}-\bar{t})^2\}^{1/2}}$$

where the sums are for i from 1 to n and where $\bar{X} = \sum_i X_{in}/n$ and $\bar{t} = \sum_i t_{in}/n$. The set $\{X_{in};\ i=1,2,\ldots,n\}$ will be referred to as an order sample; it will also be referred to simply as X, or X', where X' is the row vector $(X_{1n},X_{2n},\ldots,X_{nn})$.

Suppose the test of fit is a test of the null hypothesis H_0: X comes from the distribution $F(x;\theta)$, where θ is a vector of parameters. In many test situations, such as a test for normality, the components of θ are location and scale parameters, α and β respectively. Suppose then that the distribution is written $F(x;\alpha,\beta)$. Let $\{W_{in}, i=1,2,\ldots,n\}$ be the order statistics from a random sample of size n from $F(x;0,1)$, let E denote expectation, and let $m_{in}^{*} = E(W_{in})$ (the m_{in}^{*}'s will usually be known, or known with resonable accuracy). We now note that X_{in} can be written $\alpha+\beta W_{in}$; so if we assume that W_{in} is close to its expectation m_{in}^{*} — which will generally be true — we have the representation

$$X_{in} \sim \alpha+\beta m_{in}^{*} . \tag{1}$$

In other words, the X_{in}'s should be roughly linearly related to the m_{in}^{*}'s.

To investigate H_0 with α and β unknown it is therefore natural to use a measure of how well the linear relation (1) holds. By choosing t_{in} to be m_{in}^{*} we obtain an obvious test statistic $r(X,m^{*})$. Another procedure for testing H_0 was introduced by Shapiro and Wilk (1965). The test statistic is based on a comparison of the estimate $\hat{\beta}$ of the slope in (2), obtained by generalized least-squares, with the estimate of β obtained from the sample variance.

1.2 Test for Normality

Consider the case where we wish to test whether the ordered sample X comes from a Normal distribution, with unknown mean μ and unknown variance σ^2, written $N(\mu,\sigma^2)$. Then the W_{in} will be order statistics from $N(0,1)$; let these be called $Z_{1n} < Z_{2n} < \cdots < Z_{nn}$, let $m_{in} = E(Z_{in})$, and let m be the column vector $(m_{1n},m_{2n},\ldots,m_{nn})'$.

Let v_{ij} be the covariance between Z_{in} and Z_{jn}, that is, $v_{ij} = E(Z_{in}-m_{in})(Z_{jn}-m_{jn})$, and let V be the matrix with entries v_{ij}, $i,j=1,\ldots,n$. We now have

$$E(X_{in}) = \mu+\sigma m_{in},$$

that is, $\alpha = \mu$ and $\beta = \sigma$. The generalized least-squares estimate of β is then

$$\hat{\beta} = m'V^{-1}X/(m'V^{-1}m)$$

and the test statistics proposed by Shapiro and Wilk is

$$W = \hat{\beta}^2 R^4/(S^2C^2)$$

where $S^2 = \sum_i (X_{in}-\bar{X})^2$, $\bar{X} = \sum_i X_{in}/n$, and R^2 and C^2 are respectively the constants $m'V^{-1}m$ and $m'V^{-1}V^{-1}m$. The constants R^2 and C^2 are inserted to make $0 \leq W \leq 1$ and clearly $S^2/(n-1)$ is the sample variance estimate of β^2.

The use of this statistic is limited by the need to tabulate the coefficient vector $a' = m'V^{-1}/C$; this was done by Shapiro and Wilk (1965), using both exact and approximate methods, for $n \leq 50$. For larger values of n, Shapiro and Francia (1972) proposed the statistic

$$W' = (m'X)^2/\{(m'm)^2S^2\} ;$$

this can be seen to be equal to the correlation statistic $r_n^2(X,m)$, that is, the simple correlation between X and m. The statistic W' may be regarded as a large sample version of W because of the fact that, for large n, we may write (nonrigorously) $V^{-1}m \approx 2m$ (Stephens, 1975); then $W' \approx W$ for large n. W' is simpler than W but is still difficult to calculate because of the need to evaluate the m_{in}; also no rigorous asymptotic theory exists for W or W'.

Various authors have suggested other values for t_{in} to be used in $r_n^2(X,t)$; in particular, De Wet and Venter (1972) have proposed $t_{in} = H_{in} = \Phi^{-1}\{i/(n+1)\}$ for the normal test where the function $\Phi^{-1}(\cdot)$ is defined as follows. Let $\phi(u) = e^{-u^2/2}/(2\pi)^{1/2}$, and let

$$\Phi(z) = \int_{-\infty}^{z} \phi(u)du \ ;$$

then if $y = \Phi(z)$, $z = \Phi^{-1}(y)$.

In addition, De Wet and Venter (1972) establish the following asymptotic result:

$$2n\{1-r_n(X,H)\}-a_n \xrightarrow{\mathcal{D}} \sum_{i=3}^{\infty} (Y_i^2-1)i^{-1} \ ,$$

where X represents an ordered random sample from a normal distribution, a_n is the constant

$$a_n = E[2n\{1-r_n(X,H)\}]$$

and the Y_i are independent $N(0,1)$ variates. They provide tables for this asymptotic distribution, and also values of a_n.

It is well-known that $m_{in} \approx H_{in}$ except for extreme values and De Wet and Venter imply that the asymptotic distribution of $r_n^2(X,m)$ is the same as that of $r_n^2(X,H)$. Since W and W' have become well established as tests of normality, and have good power properties, it would be valuable to have rigorous asymptotic theory.

In this article we first give, in Theorem 1, approximations for m_{in}, together with expressions for the error in the approximations. These should have some independent interest, but in Theorem 2 they are used to show that $r_n^2(X,m)$ does indeed have the same asymptotic properties as $r_n^2(X,H)$. The theorems are given in Section 2, and the proofs appear in Sections 3 and 4.

1.3 Notation

In addition to the notation established above, it will be convenient to list further definitions which will be required in the later sections. From now on, X will refer to an ordered random sample from $N(\mu,\sigma^2)$, and vector Z, with components Z_{in}, $i=1,\ldots,n$, will be a similar vector from $N(0,1)$. Then let U_i, $i=1,\ldots,n$ be defined by $U_i \equiv \Phi(Z_{in})$; vector U with components U_i, $i=1,\ldots,n$ will then be a vector of order statistics from the uniform distribution with limits 0,1, written $U(0,1)$; the dropping of the second subscript n in component U_i is done to facilitate the printing of the algebraic calculations in Section 3 and 4, and related quantities will likewise be simplified in notation. Thus we define $V_i = -\log(U_i)$; note that $V_1 > V_2 > \cdots > V_n$ are order statistics from a standard exponential distribution. Let $\psi(\cdot)$ be the function defined by $Z_{in} = \Phi^{-1}(U_i) = \Phi^{-1}\{\exp(-V_i)\} = \psi(V_i)$, $i=1,\ldots,n$.

Set $Y(u) \equiv \psi''(\log u^{-1})$, and let $s_i = E(V_i) = \Sigma_{\nu=i}^{n} (1/\nu)$; also define $p_i^0 = e^{-s_i}$.

The end of a proof will be marked by $\|$.

2. TWO THEOREMS

2.1 An Approximation for m_{in}.

Let Z_{in}, m_{in}, V_i and s_i, for $i=1,\ldots,n$, be as defined in Section 1.3. We have

THEOREM 1: The mean m_{in} of Z_{in} is given by

$$m_{in} = \Phi^{-1}\{\exp(-s_i)\} + R_{in}, \quad 1 \leq i \leq [\tfrac{n+1}{2}]$$

where

$$|R_{in}| \leq Ci^{-1}\{\log(n/i)\}^{-3/2}$$

and C is a constant independent of i and n.

REMARKS. (1) $m_{in} = -m_{jn}$, where $j = n+1-i$, so the theorem covers $1 \leq i \leq n$.

(2) For i far enough away from 1, $e^{-s_i} \simeq i/(n+1)$; the theorem then gives $m_{in} \approx \Phi^{-1}\{i/(n+1)\} = H_{in}$ above, a familiar approximation for m_{in}. Blom (1958) introduced the idea of using exponential order statistics, although he expanded Z_{in} about $\log\{i/(n+1)\}$ and restricted his attention to the case of i fixed as $n \to \infty$.

(3) A crude expression for the error term when H_{in} is used to approximate m_{in} was stated by David and Johnson (1956) but this requires i/n to be bounded away from both 0 and 1. Lemma 6 below shows that

when H_{in} is used to approximate m_{in} the error term is of order $i^{-1}\{\log(n/i)\}^{-1/2}$, $1 \leq i \leq [(n+1)/2]$. Thus $\Phi^{-1}(\exp(-s_i))$ provides a better approximation, at least asymptotically.

2.2 The next theorem establishes the equivalence of the asymptotic distributions for $r_n(Z,m)$ and $r_n(Z,H)$.

<u>THEOREM 2</u>: <u>With</u> $r_n(Z,m)$ <u>and</u> $r_n(Z,H)$ <u>defined as in Section 1.1</u>,

$$n\{r_n(Z,m) - r_n(Z,H)\} \to 0$$

<u>in probability</u>.

Since, under H_0, $Z_{in} = (X_{in}-\mu)/\sigma$, it follows that $r_n(X,m) \equiv r_n(Z,m)$ and $r_n(X,H) \equiv r_n(Z,H)$. Thus Theorem 2 asserts that the Shapiro-Francia and the De Wet-Venter correlation statistics have the same null asymptotic distributions.

In the next two sections we give the proofs of Theorems 1 and 2.

3. PROOF OF THEOREM 1

3.1 Throughout the section we suppose $1 \le i \le [(n+1)/2]$. In this paragraph the first steps in the proof are given. These steps motivate a series of lemmas which are required to complete the proof. To begin, we expand Z_{in} about $s_i = \Sigma_{\nu=i}^{n} 1/\nu$. This gives, using the notation $Z_{in} = \Phi^{-1}(e^{-V_i}) \equiv \Psi(V_i)$,

$$Z_{in} = \Psi(s_i) + (V_i - s_i)\Psi'(s_i) + \frac{1}{2}(V_i - s_i)^2 \Psi''(\theta_i)$$

where θ_i lies between s_i and V_i. Taking expectations gives

$$m_{in} = E(Z_{in}) = \Psi(s_i) + E\,\frac{1}{2}(V_i - s_i)^2 \Psi''(\theta_i) .$$

We need to evaluate $E(V_i - s_i)^2 \Psi''(\theta_i)$. Recall that $U_i = \exp(-V_i)$, and let $g_{in}(\cdot)$ be the density of U_i; we have

$$g_{in}(u) = n\binom{n-1}{i-1} u^{i-1}(1-u)^{n-i}, \; 0 < u < 1 .$$

Also with $p_i^0 = \exp(-s_i)$, we define h_{in} by:

$$h_{in} = \begin{cases} p_i^0[1+\lambda(i-1)^{-1/2}\log n] & \text{for } i > 1+\lambda^2 \log n , \\ p_i^0(1+\lambda^3 i^{-1}\log n) & \text{for } 1 \le i \le 1+\lambda^2 \log n , \end{cases}$$

with $\lambda = 10$. Then

$$I = E(V_i - s_i)^2 \Psi''(\theta_i) = \int_0^1 \gamma(u)du = I_1 + I_2 ,$$

where $\gamma(u) = (\log u^{-1} - s_i)^2 \Psi''(\theta_i) g_{in}(u)$, and

$$(2) \qquad I_1 = \int_0^{h_{in}} \gamma(u)du \quad \text{and} \quad I_2 = \int_{h_{in}}^1 \gamma(u)du .$$

I must now be evaluated by evaluating I_1 and I_2. To do this, a series of lemmas is needed. A key result in the proof of Theorem 1 is Lemma 1.

<u>LEMMA 1</u>. <u>The function</u> $Y(u) = \Psi''(\log u^{-1})$ <u>is monotonic increasing and positive for</u> $0 \leq u \leq 0.5$; <u>also</u>

$$(3) \qquad \begin{aligned} Y(u) &< (|\Phi^{-1}(u)|)^{-3} && \text{for } \Phi^{-1}(u) < 0 ; \\ Y(u) &< 16 && \text{for } \Phi^{-1}(u) \leq 1 . \end{aligned}$$

<u>COROLLARY</u>:

$$Y(u) \leq 32(1+|\Phi^{-1}(u)|^3)^{-1} \ \underline{\text{for}} \ \Phi^{-1}(u) \leq 1, \ \underline{\text{that is, for}} \ 0 \leq u \leq \Phi(1) \approx .84.$$

<u>PROOF OF LEMMA 1</u>. It is easily shown that $Y(u) = A(u)B(u)$ where

$$A(u) = u\{\phi(\Phi^{-1}(u)\}^{-1} \quad \text{and} \quad B(u) = 1+u\Phi^{-1}(u)[\phi\{\Phi^{-1}(u)\}]^{-1} .$$

We therefore prove that $Y(u)$ is monotonic by proving that both $A(u)$ and $B(u)$ are monotonic.

<u>Monotonicity of A(u)</u>. Let $u = \Phi(x)$, and let $A(u) \equiv A_1(x)$. Since $du/dx > 0$, $A(u)$ is monotonic increasing if $A_1'(x) > 0$. We have $A_1(x) = \Phi(x)/\phi(x)$; thus $A_1'(x) = C(x)/\phi(x)$, where $C(x) = \phi(x)+x\Phi(x)$. It is easily shown that $\lim_{x\to-\infty} C(x) = 0$; also $C'(x) = \Phi(x)$, using the fact that $\phi'(x) = -x\phi(x)$; thus $C'(x) > 0$, and $C(x)$ is positive and monotonic increasing for all x. Hence $A_1'(x)$ is positive, so $A(u)$ is monotonic increasing for all x.

<u>Monotonicity of B(u)</u>. Let $B(u) \equiv B_1(x)$. Define the function $D(x) = (1+x^2)\Phi(x)+x\phi(x)$. We have $D'(x) = 2C(x) > 0$; also $\lim_{x\to-\infty} D(x) = 0$, so $D(x)$ is positive and monotonic increasing for all x. Now $B_1(x) = 1+x\Phi(x)/\phi(x) = C(x)/\phi(x)$; then $B_1'(x) = \{C'(x)+xC(x)\}/\phi(x) = \{(1+x^2)\Phi(x) + x\phi(x)\}/\phi(x) = D(x)/\phi(x)$. Thus $B_1'(x) > 0$; hence $B'(u) > 0$ and so $B(u)$ is monotonically increasing for all x.

It is well-known (see, for example, Renyi (1970, p. 164) that for $x < 0$,

$$(4) \qquad 1-x^{-2} < |x|\Phi(x)/\phi(x) < 1 \; ;$$

thus, for $x < 0$, $A_1(x) \equiv \Phi(x)/\phi(x) < 1/|x|$, and $B_1(x) \equiv 1+x\Phi(x)/\phi(x) < x^{-2}$; hence $Y(u) \equiv B_1(x)A_1(x) < 1/|x|^3 \equiv |\Phi^{-1}(u)|^{-3}$ for $u < 0.5$, that is, for $x < 0$. For $0.5 < u < \Phi(1)$, that is, for $0 < x < 1$, $Y(u)$ is increasing, so for the second inequality we evaluate $Y(u)$ at $u = \Phi(1)$; the value is approximately 15.6. This completes the proof of Lemma 1.

LEMMA 2: There is a K, independent of i, such that for all $n \geq K$,

$$h_{in} \leq \Phi(1),\ 1 \leq i \leq [\tfrac{1}{2}(n+1)] .$$

PROOF OF LEMMA 2. We have

$$\sum_{\nu=i}^{n-1} \nu^{-1} < \int_{i-1}^{n-1} x^{-1}dx = \log(\frac{n-1}{i-1}) < \sum_{\nu=i-1}^{n-2} \nu^{-1} ;$$

so

$$e^{-1/n}(\frac{i-1}{n-1}) < p_i^0 = e^{-s_i} < \frac{i}{n+1} < \frac{i}{n} . \tag{5}$$

From the definition of h_{in} the lemma follows. ||

LEMMA 3: For a constant c_0 $(0 < c_0 < 1)$, there is a constant $\gamma(c_0)$ such that for $0 < u < \gamma(c_0) < \frac{1}{2}$,

$$-\{-\log(2\pi u^2)\}^{1/2} < \Phi^{-1}(u) < -\{-c_0\log(2\pi u^2)\}^{1/2} .$$

PROOF OF LEMMA 3. By (4) we have

$$\phi(z)|z|^{-1}(1-z^{-2}) < \Phi(z) < \phi(z)|z|^{-1} ;$$

set $v = -\log(2\pi u^2)$ and $z = -(c_0 v)^{1/2}$; then

$$u^{-1}\Phi\{-(c_0v)^{1/2}\} > \frac{\phi\{-(c_0v)^{1/2}\}}{u(c_0v)^{1/2}}\{1-(c_0v)^{-1}\}$$

$$= \{1-(c_0v)^{-1}\}/\{(u\sqrt{2\pi})^{1-c_0}(c_0v)^{1/2}\}$$

$$= h(u,c_0), \quad \text{say} .$$

Now $h(u,c_0) \to \infty$ as $u \downarrow 0$ so there is a constant $\gamma_0(c_0)$ such that $h(u,c_0) > 1$ for $0 \le u < \gamma_0(c_0)$. Similarly

$$u^{-1}\Phi(-v^{1/2}) < u^{-1}\phi(-v^{1/2})/v^{1/2}$$

$$= v^{-1/2}$$

$$< 1, \quad \text{provided} \quad u < (2\pi e)^{-1/2} .$$

Thus, if $\gamma(c_0) = \min\{\gamma_0(c_0),(2\pi e)^{-1/2}\}$, we have for $0 \le u \le \gamma_0(c_0)$, $u^{-1}\Phi(-v^{1/2}) < 1$ and $u^{-1}\Phi\{-(c_0v)^{1/2}\} > 1$. Lemma 3 follows at once. ||

LEMMA 4: *With p_i^0 and h_{in} defined as in Section 3.1, there is a K' such that for $n \ge K'$*

$$|\Phi^{-1}(h_{in})| \ge C|\Phi^{-1}(p_i^0)|, \quad 1 \le i \le [\tfrac{n+1}{2}]$$

where C is independent of i and n.

PROOF OF LEMMA 4. By Lemma 3, when $h_{in} < \gamma(c_0)$ with c_0 fixed $(0 < c_0 < 1)$ but otherwise arbitrary — for example, c_0 can be 0.5 –

$$|\Phi^{-1}(h_{in})| \geq \{-c_0 \log(2\pi h_{in}^2)\}^{1/2}$$

$$> \begin{cases} c_0^{1/2}[-2\log\{p_i^0(2\lambda^3)i^{-1}(\log n)\sqrt{2\pi}\}]^{1/2}, 1 \leq i \leq 1+\lambda^2 \log n \\ c_0^{1/2}(-2\log[p_i^0\{1+\lambda(i-1)^{-1/2}(\log n)^{1/2}\}\sqrt{2\pi}])^{1/2}, \ 1+\lambda^2 \log n < i \leq \dfrac{n}{4\sqrt{2\pi}} . \end{cases}$$

Using (5) we obtain

$$|\Phi^{-1}(h_{in})| > \begin{cases} c_0^{1/2}|\Phi^{-1}(p_i^0)|[1 - \dfrac{\log\{(2\lambda^3)\log n\}}{\log\{n(1+\lambda^2\log n)^{-1}(2\pi)^{-1/2}\}}]^{1/2}, \ 1 < i < 1+\lambda^2 \log n \\ \\ c_0^{1/2}|\Phi^{-1}(p_i^0)|(1-\log 2/\log 4)^{1/2}, \ 1+\lambda^2 \log n < i \leq n(32\pi)^{-1/2} \end{cases}$$

and so

$$|\Phi^{-1}(h_{in})| > c_1|\Phi^{-1}(p_i^0)| \quad \text{for} \quad 1 \leq i \leq n(32\pi)^{-1/2} \quad \text{and } h_{in} < \gamma(c_0) .$$

When $n > K$, by Lemma 2, $h_{in} \leq \Phi(1)$, and by (5)

$$n(32\pi)^{-1/2} < i \leq [\tfrac{1}{2}(n+1)] \Rightarrow \tfrac{1}{2}(32\pi)^{-1/2} < p_i^0 \leq \tfrac{1}{2}$$

$$\Rightarrow h_{in} > \frac{1}{8\sqrt{2\pi}} .$$

For $\gamma(c_0) \leq h_{in} \leq \Phi(1)$, there is a c_2, independent of i and n, such that $c_2 < p_i^0 \leq \frac{1}{2}$. From these results taken together, there must be a c_3 and c_4, both independent of i and n, such that with $n > K$,

$$|\Phi^{-1}(h_{in})| > \begin{cases} c_3|\Phi^{-1}(p_i^0)| & \text{for } (32\pi)^{-1/2}n < i \le [\frac{1}{2}(n+1)] , \\ c_4|\Phi^{-1}(p_i^0)| & \text{for } 1 \le i \le (32\pi)^{-1/2}n, \ \gamma(c_0) \le h_{in} \le \Phi(1) . \end{cases}$$

This completes the proof of Lemma 4. ||

LEMMA 5: For $1 \le i \le [\frac{1}{2}(n+1)]$ and k a fixed positive integer,

$$\int_{\Phi(1)}^{1} (1-u)^{-k} g_{in}(u)du = o(n^{-m})$$

where m is an arbitrary positive real number.

PROOF OF LEMMA 5. Let

$$f_{in}(u) = \frac{(n-1)^i}{(i-1)!} u^{i-1} e^{-(n-1)u}, \quad u \ge 0 .$$

Lemma 3 of Stigler (1969, p. 774) gives the result that for any $\varepsilon > 0$, there is an M depending only on ε such that $g_{in}(u) \le Mf_{in}(u)$ for all $u \ge 0$ and $i \le (1-\varepsilon)n$. Therefore

$$\int_{\Phi(1)}^{1} (1-u)^{-k} g_{in}(u)du = \frac{n(n-1)\dots(n-k+1)}{(n-i)\dots(n-i-k+1)} \int_{\Phi(1)}^{1} g_{i(n-k)}(u)du$$

$$< 3^k M \int_{\Phi(1)}^{1} f_{i(n-k)}(u)du \quad \text{for} \quad n > 4k$$

$$< C \exp\{-\frac{1}{2} n\Phi(1)\} \int_{\Phi(1)}^{1} \frac{(n-k-1)^i}{(i-1)!} u^{i-1} e^{-u(n-k-1)/2} du$$

$$< C \exp\{-\frac{1}{2} n\Phi(1)\} 2^i$$

$$= o(n^{-m}) \quad \text{as} \quad 2^i \le e^{\{(n+1)\log 2\}/2} \le e^{.4(n+1)}, \ e^{-n\Phi(1)/2} < e^{-.42n} .$$

This completes the proof of Lemma 5. $\|$

3.3 We continue with the proof of Theorem 1. We assume $n > \max(K,K')$. From (2) in Section 3.1 we have

$$I_1 = \int_0^{h_{in}} (\log u^{-1}-s_i)^2\Psi''(\theta_i)g_{in}(u)du$$

$$\leq C[1+|\Phi^{-1}(p_i^0)|^3]^{-1} \sum_{\nu=1}^{n} \nu^{-2}$$

using Lemmas 1 and 4, and the fact that $E(V_i-s_i)^2 = \Sigma_i^n \nu^{-2}$. Also $I_2 = I_{21}+I_{22}$ where

$$I_{21} = \int_{h_{in}}^{\Phi(1)} (\log u^{-1}-s_i)^2\Psi''(\theta_i)g_{in}(u)du$$

and

$$I_{22} = \int_{\Phi(1)}^{1} (\log u^{-1}-s_i)^2\Psi''(\theta_i)g_{in}(u)du \ .$$

By Lemma 1, we can write

(6) $$I_{21} \leq 16s_i^2 \int_{h_{in}}^{\Phi(1)} g_{in}(u)du \ ,$$

using the fact that $\log u^{-1} < s_i$ throughout the range of the integral. To evaluate the integral in (6) we use an argument given in Lemma 4 of Stigler (1969). (The lemma as stated, however, has a condition missing. Using the notation of Stigler (1969), let $(i-1) < \lambda^2\log n$, $h(u) = 1$, $k = 0$, and consider A_n',

$$\int_{\gamma_n'}^{\infty} f_{in}(u)du = \sum_{j=0}^{i-1} e^{-(n-1)\gamma_n'}((n-1)\gamma_n')^j/j!$$

$$\geq e^{-(n-1)\gamma_n'} \simeq e^{-\lambda(i-1)\log n)^{1/2}} .$$

Suppose $i = \log(\log n)$; the right hand side will not then be $o(n^{-m})$ for any $m > 0$. The condition $b_n/\log n \to \infty$ is also necessary. It is to avoid this difficulty that we define h_{in} over two regions for i.) By (5), for $i > 1+\lambda^2\log n$, we have

$$h_{in} = p_i^0[1+\lambda\{\log(n)/(i-1)\}^{1/2}]$$

$$> \frac{(i-1)}{(n-1)} e^{-1/n}[1+\lambda\{\log(n)/(i-1)\}^{1/2}]$$

$$> \frac{(i-1)}{n} [1+\lambda\{\log(n)/(i-1)\}^{1/2}] = h_{in}^* \quad \text{for} \quad n > K'',$$

where K'' is an appropriate constant.

For $1 \leq i \leq 1+\lambda^2\log(n)$ we have

$$h_{in} = p_i^0(1+\lambda^3 i^{-1}\log n) > .45\lambda^3(n-1)^{-1}\log n \quad \text{for} \quad n > K''' ,$$

where K''' is an appropriate constant.

using the fact that $p_i^0 > (i-\frac{1}{2})/(n+\frac{1}{2})$. Thus for $i > 1 + \lambda^2\log n$, noting that $\lambda = 10$ and $f_{in}(u)$ is decreasing for $u > (i-1)/(n-1)$, we find

$$\int_{h_{in}}^{\Phi(1)} g_{in}(u)du \le M \int_{h_{in}^*}^{1} f_{in}(u)du$$

$$< Cn[1+\lambda\{\log(n)/(i-1)\}^{1/2}]^{i-1}e^{-\lambda\{(i-1)\log n\}^{1/2}}$$

$$= O(n^{-2}) ,$$

since $(1+a/r)^r \le \exp[a-\{a^2/(12r)\}]$, $0 \le a \le r$.

For $1 \le i \le 1+\lambda^2\log n$, we have, using $w_n = 0.45\lambda^3(n-1)^{-1}\log n$,

$$\int_{h_{in}}^{\Phi(1)} g_{in}(u)du \le M \int_{h_{in}}^{\Phi(1)} f_{in}(u)du$$

$$\le M \int_{w_n}^{\infty} f_{in}(u)du$$

$$\le M\{\exp(-.225\lambda^3\log n)\}2^i \int_0^{\infty} \frac{\{(n-1)/2\}^i}{(i-1)!} u^{i-1}e^{-u(n-1)/2}du$$

$$\le M \exp[-\{.225\lambda^3-(1+\lambda^2)\log 2\}\log n]$$

$$= o(n^{-2}) .$$

Hence

$$I_{21} = o(n^{-1}\log n)^2$$

and, writing u^* for $e^{-\theta_i}$,

$$I_{22} \le s_i^2 \int_{\Phi(1)}^{1} \frac{u^*(1-u^*)}{\phi\{\Phi^{-1}(u)^*\}} [(1-u^*) + \frac{u^*(1-u^*)\Phi^{-1}(u^*)}{\phi\{\Phi^{-1}(u^*)\}}](1-u)^{-2} g_{in}(u)du$$

since $p_i^0 < u^* < u$ and hence $(1-u^*)^{-1} < (1-u)^{-1}$.

Note that (4) holds for $x > 0$ with $\Phi(x)$ replaced by $1-\Phi(x)$. We apply (4) with $x = \Phi(u^*)$; this allows us to bound the term in I_{22} involving u^* irrespective of whether $u^* < \frac{1}{2}$ or $u^* > \frac{1}{2}$. Thus by Lemma 5, we have

$$I_{22} \le C\, s_i^2\, n^{-3} .$$

Finally by equation (5) and Lemma 3, we obtain

$$|\Phi^{-1}(p_i^0)| > |\Phi^{-1}(\frac{i}{n})| > C\{\log(\frac{n}{i})\}^{1/2},\ 1 \le i \le [\frac{1}{2}(n+1)] .$$

This completes the proof of Theorem 1. ||

4. PROOF OF THEOREM 2

4.1 Another series of lemmas is needed for the proof of Theorem 2. Let $p_i = i/(n+1)$ and for a vector $a = (a_1,\ldots,a_n)'$, define the norm $\|\cdot\|$ by

$$\|a\| = (\sum_1^n a_i^2)^{1/2} .$$

<u>LEMMA 6</u>: <u>The following inequalities hold</u>:

$$|m_{in}-H_{in}| \le ci^{-1}(1+H_{in})^{-1} \le c'i^{-1}\{\log(n/i)\}^{-1/2}, \ \underline{\text{for}}\ 1 \le i \le [\tfrac{1}{2}(n+1)]$$

<u>where</u> c, c' <u>are constants, independent of</u> i <u>and</u> n.

<u>PROOF OF LEMMA 6</u>: By Theorem 1,

$$|m_{in}-H_{in}| = |\Phi^{-1}(p_i^0)-\Phi^{-1}(p_i)+R_{in}|$$

and

$$\Phi^{-1}(p_i^0)-\Phi^{-1}(p_i) = [s_i-\log\{(n+1)i^{-1}\}]p_i/\phi\{\Phi^{-1}(p_i)\} +$$
$$\tfrac{1}{2}[s_i-\log\{(n+1)i^{-1}\}]^2\Psi''(\theta^*)$$

where

$$\log\{(n+1)/i\} \le \theta^* \le s_i .$$

Now

$$s_i > \log\{(n+1)/i\}$$

and

$$s_i = \sum_{\nu=i+1}^{n+1} \nu^{-1}+i^{-1}-(n+1)^{-1} < \int_i^{n+1} x^{-1}dx+i^{-1}-(n+1)^{-1} ;$$

so

$$0 < s_i - \log\{(n+1)/i\} < i^{-1} .$$

By (4) and the fact that $\Psi''(\theta)$ is decreasing (see Lemma 1), we have

$$|m_{in}-H_{in}| \le ci^{-1}(1+H_{in})^{-1}$$

$$\le c'i^{-1}\{\log(n/i)\}^{-1/2} \text{ by Lemma 3.}$$

This completes the proof of Lemma 6. ||

<u>LEMMA 7</u>: <u>The norm of vector</u> m-H <u>satisfies</u>

$\|m-H\| < c(\log n)^{-1/2}$, <u>where</u> c <u>is a constant independent of</u> n .

<u>PROOF OF LEMMA 7</u>: From Lemma 6 we can write

$$\|m-H\|^2 = 2\sum_{i=1}^{[(n+1)/2]} (m_{in}-H_{in})^2$$

$$\leq c' \sum_{i=1}^{[(n+1)/2]} i^{-2}\{\log(n/i)\}^{-1} < c''n^{-1}\int_2^n (\log x)^{-1}dx$$

$$\leq c(\log n)^{-1} ,$$

where c, c', c'' are constants independent of n. ||

<u>LEMMA 8</u>: <u>The norm of vector</u> H <u>satisfies</u>

$$n-5\log n \leq \|H\|^2 \leq n+1 \quad \underline{\text{for}} \quad n \geq 24 .$$

<u>PROOF OF LEMMA 8</u>: For the second inequality

$$\|H\|^2 \leq (n+1)\int_0^1 [\Phi^{-1}(u)]^2 du = n+1 .$$

For the first inequality

$$\|H\|^2 = 2\sum_{1}^{[(n+1)/2]} [\Phi^{-1}\{i/(n+1)\}]^2$$

$$\geq 2(n+1)\int_{(n+1)^{-1}}^{1/2} [\Phi^{-1}(u)]^2 du$$

$$= (n+1)[1-2\int_{-\infty}^{\Phi^{-1}(1/(n+1))} u^2\phi(u)du]$$

$$= (n+1)[1-2|\Phi^{-1}\{1/(n+1)\}|\phi[\Phi^{-1}\{1/(n+1)\}]-2(n+1)^{-1}]$$

$$\geq n-3-2[\Phi\{1/(n+1)\}]^2-2[[\Phi^{-1}\{1/(n+1)\}]^2-1)^{-1} \text{ by (4)}$$

$$\geq n-5\log n \text{ by Lemma 3 and the fact that } |\Phi^{-1}\{1/(n+1)\}| > 1.73$$

for $n \geq 24$. ||

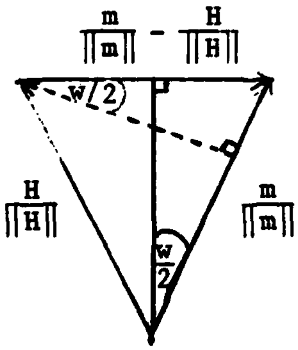
<u>LEMMA 9</u>: <u>The normalised vectors for</u> m <u>and</u> H <u>satisfy</u>

$$\left\|\frac{m}{\|m\|} - \frac{H}{\|H\|}\right\| \leq \sqrt{2}\,\|H\|^{-1}\,\|m-H\| .$$

<u>COROLLARY</u>:

$$0 < \|m\|\,\|H\| - \sum_1^n H_{in}m_{in} \leq \|m-H\|^2\,\|m\|/\|H\| .$$

<u>PROOF OF LEMMA 9 AND COROLLARY</u>: This is the same as that given by Sarkadi (1975, p. 447) for a similar result. It depends on the fact that if $\{a_i\}$ and $\{b_i\}$ are both increasing sequences of n real numbers, then $\Sigma_1^n a_ib_i \geq 0$ if either $\Sigma_1^n a_i = 0$ or $\Sigma_1^n b_i = 0$. In particular $\Sigma_1^n m_iH_i \geq 0$.



Then in the sketch, with w representing the angle between m and H ,

$$\left\|\frac{m}{\|m\|} - \frac{H}{\|H\|}\right\| \cos(w/2)$$

$\leq$ "length of any line from **top** of $\frac{H}{\|H\|}$ to a point on m"

$$\leq \left\|\frac{H}{\|H\|} - \frac{m}{\|H\|}\right\| = \|H\|^{-1}\,\|m-H\| .$$

Also, geometrically, the fact that $\Sigma_1^n m_i H_i \geq 0$ means that $w \leq \pi/2$ and $\cos w \geq 2^{\frac{1}{2}}$. More formally,

$$\cos(w/2) = \frac{1}{2} \left\| \frac{m}{\|m\|} + \frac{H}{\|H\|} \right\|$$

$$= 2^{-1/2} \left(1 + \frac{\Sigma_1^n m_i H_i}{\|m\| \, \|H\|} \right)^{1/2}$$

$$\geq 2^{-1/2} .$$ Lemma 9 follows at once $\|$.

<u>LEMMA 10</u>: <u>The components of</u> m <u>and</u> H <u>satisfy</u>

$$\sum_{i=1}^{n} |H_{in}| \, |m_{in} - H_{in}| \leq c \log(n) .$$

<u>PROOF OF LEMMA 10</u>: By Lemma 6

$$\sum_{1}^{n} |H_{in}| \, |m_{in} - H_{in}| \leq c \sum_{1}^{[(n+1)/2]} |H_{in}| i^{-1} (1 + |H_{in}|)^{-1} \leq c' \sum_{1}^{[(n+1)/2]} i^{-1} . \; \|$$

<u>LEMMA 11</u>: <u>The following inequality holds</u>:

$$E|Z_{in} - \Phi^{-1}(p_i^0)| \leq c\{i \log(n/i)\}^{-1/2}, \quad 1 \leq i \leq [\tfrac{1}{2}(n+1)]$$

<u>where</u> c <u>is a constant independent of</u> i <u>and</u> n.

<u>PROOF OF LEMMA 11</u>: The formal proof of this Lemma follows the same steps as those used to establish Theorem 1. In this case we expand only to the first term

$$Z_{in} = \Phi^{-1}(p_i^0) + (V_i - s_i)\Psi'(\theta_i)$$

where $\Psi'(v) = -e^{-v}/\phi(\Phi^{-1}(e^{-v}))$ and θ_i is between V_i and s_i. This yields

$$E|Z_{in} - \Phi^{-1}(p_i^0)| = E|V_i - s_i||\Psi'(\theta_i)| \ .$$

By Lemma 1, $\Psi'(v) = -e^{-v}\{\phi(\Phi^{-1}(e^{-v}))\}^{-1}$ is increasing in v and is negative. Hence, $|\Psi'(\theta)|$ is decreasing as θ increases and by (4)

$$|\Psi'(\log u^{-1})| \leq \begin{cases} |\Phi^{-1}(u)|^{-1} & \text{for } \Phi^{-1}(u) < -1 \\ \\ 4 & \text{for } |\Phi^{-1}(u)| \leq 1 \ . \end{cases}$$

Therefore

$$|\Psi'(\log u^{-1})| < 8\{1 + |\Phi^{-1}(u)|\},\ \Phi^{-1}(u) \leq 1,\ \text{that is, for } 0 \leq u \leq \Phi(1) \ .$$

Now

$$E\{|V_i - s_i||\Psi'(\theta_i)|\} = \int_0^{h_{in}} + \int_{h_{in}}^{\Phi(1)} + \int_{\Phi(1)}^1 |\log u^{-1} - s_i||\Psi'(\theta_i)|g_{in}(u)du \ .$$

These integrals can be evaluated exactly as before, yielding the upper bound

$$c' E\{|V_{in}-s_i|\}/\{1+|\Phi^{-1}(p_i^0)|\} + O(n^{-2}) \ ,$$

which in turn is bounded by

$$c\{i \log(n/i)\}^{-1/2} \ . \ \|$$

<u>LEMMA 12</u>: <u>The norm of vector</u> m <u>satisfies</u>

$$0 \le 1-n^{-1}\|m\|^2 \le cn^{-1}\log n \quad \underline{\text{for all}} \quad n \ ,$$

<u>where</u> c <u>is a constant independent of</u> n.

<u>PROOF OF LEMMA 12</u>:

$$(E\{|Z_{in}|\})^2 \le E\{Z_{in}^2\} \ .$$

So

$$n^{-1}\|m\|^2 \le n^{-1}\sum_1^n E\{Z_{in}^2\} = 1 \ .$$

Also

$$n^{-1}\|m\|^2 = n^{-1}\|H\|^2 + n^{-1}\|m-H\|^2 + 2n^{-1}\sum_1^n H_{in}(m_{in}-H_{in})$$

$$\ge 1-6n^{-1}\log n - c(n\log n)^{-1} - c'n^{-1}\log n \ ,$$

By Lemmas 7, 8, and 10. $\|$

LEMMA 13: The norms of vectors m and H satisfy

$$n^{1/2}\left| \|m\|^{-1} - \|H\|^{-1} \right| \le cn^{-1}\log n ,$$

where c is a constant independent of n.

PROOF OF LEMMA 13: We have

$$\left| \|m\| - \|H\| \right| = \left| \sum_1^n (m_{in}-H_{in})^2 + 2 \sum_1^n H_{in}(m_{in}-H_{in}) \right| / (\|m\|+\|H\|)$$

$$\le cn^{-1/2}\log n \text{ by Lemmas 7, 8, 10 and 12.} \quad \|$$

4.2. We can now turn to the major proof.

PROOF OF THEOREM 2: We have

$$n\{r_n(Z,m)-r_n(Z,H)\} = S_n^{-1} \sum_1^n Z_{in}[m_{in}M_n^{-1}-H_{in}K_n^{-1}]$$

where

$$M_n^2 = n^{-1} \sum_1^n m_{in}^2, \quad K_n^2 = n^{-1} \sum_1^n H_{in}^2, \quad S_n^2 = n^{-1} \sum_{i=1}^n (Z_{in}-\overline{Z})^2$$

and

$$\overline{Z} = n^{-1} \sum_{i=1}^n Z_{in} .$$

As $S_n^2 \xrightarrow{a.s.} 1$, $n\{r_n(Z,m)-r_n(Z,H)\} \xrightarrow{p} 0$ if and only if

$$A_n = \sum_{i=1}^{n} Z_{in}(m_{in}M_n^{-1}-H_{in}K_n^{-1}) \xrightarrow{p} 0 .$$

Now

$$A_n = \sum_{i=1}^{n} \{Z_{in}-\Phi^{-1}(p_i^0)\}\{m_{in}M_n^{-1}-H_{in}K_n^{-1}\} + \sum_{1}^{n} \{\Phi^{-1}(p_i^0)-H_{in}\}\{m_{in}M_n^{-1}-H_{in}K_n^{-1}\}$$

$$+ M_n^{-1} (\sum_{1}^{n} H_{in}m_{in}-\|m\| \|H\|)$$

$$= B_n + C_n + D_n, \quad \text{say} .$$

Using Markov's inequality $B_n \xrightarrow{p} 0$ if $E|B_n| \to 0$. By Lemmas 6, 11, 12 and 13, we have

$$E|B_n| \leq \sum_{i=1}^{n} \{E|Z_{in}-\Phi^{-1}(p_i^0)|\}(M_n^{-1}|m_{in}-H_{in}|+|H_{in}||M_n^{-1}-K_n^{-1}|)$$

$$< c \sum_{i=1}^{[(n+1)/2]} \{i \log(n/i)\}^{-1/2}[ci^{-1}\{\log(n/i)\}^{-1/2}+\{\log(n/i)\}^{1/2} n^{-1}\log n]$$

$$< c(\log n)^{-1} \to 0 \quad \text{as} \quad n \to \infty .$$

By the proof of Lemma 6,

$$C_n \leq c \sum_{1}^{[(n+1)/2]} i^{-1}\{\log(n/i)\}^{-1/2}[c'i^{-1}\{\log(n/i)\}^{-1/2}$$

$$+ \{\log(n/i)\}^{1/2}n^{-1}\log n]$$

$$\leq c(\log n)^{-1} .$$

Finally, by the corollary to Lemma 9, and Lemmas 7, 8, and 12,

$$D_n \to 0 \quad \text{as} \quad n \to \infty .$$

This completes the proof of Theorem 2. $\|$

REFERENCES

[1] Blom, G. (1958). Statistical Estimates and Transformed Beta-Variables. New York, John Wiley and Sons, Inc.

[2] David, F.N. and Johnson, N.L. (1956). Statistical Treatment of Censored Data I. Fundamental Formulae. Biometrika, 41, 228-240.

[3] DeWet, T. and Venter, J.H. (1972). Asymptotic Distributions of Certain Test Criteria of Normality. S. Afr. Statist. J., 6, 135-149.

[4] Renyi, A. (1970). Probability Theory. North-Holland, Amsterdam.

[5] Sarkadi, K. (1975). The Consistency of the Shapiro-Francia Test. Biometrika, 62, 2, 445-450.

[6] Shapiro, S.S. and Francia, R.S. (1972). An Approximate Analysis of Variance Test for Normality. J. American Statist. Assoc., 67, (March 1972) 337, 215-216.

[7] Shapiro, S.S. and Wilk, M.B. (1965). An Analysis of Variance Test for Normality. Biometrika, 42, 591-611.

[8] Stephens, M.A. (1975). Asymptotic Properties for Covariance Matrices of Order Statistics. Biometrika, 62, 1, 23-28.

[9] Stigler, S.M. (1969). Linear Functions of Order Statistics. Ann. Math. Statist., 40, 3, 770-788.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 342	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Approximations For Expected Values Of Normal Order Statistics With An Application To Goodness-Of-Fit		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) N. Fotopoulos, J. Leslie and M. Stephens		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0475
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 411SP		12. REPORT DATE March 6, 1984
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Correlation tests for normality; de Wet-Venter test for normality; Normal order statistics; Shapiro-Francia test; Shapiro-Wilk test; Tests of normality

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this article we consider new approximations for expected values of standard normal order statistics, and uses ~~them~~ to give asymptotic theory for the Shapiro-Francia (1972) statistic for a test of normality.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

5-84

DTIC